DOCUMENT RESUME

ED 463 526 CS 510 851

AUTHOR Rhodes-Kline, Anne K.
TITLE Testing the Whole Class: What Impact Does It Have?
INSTITUTION Maine Univ., Orono. Center for Early Literacy.
PUB DATE 1997-07-00
NOTE 10p.
PUB TYPE Reports - Research (143)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Chi Square; Classroom Research; Data Collection; Early Intervention; Grade 1; *High Risk Students; Kindergarten; *Literacy; Primary Education; *Testing
IDENTIFIERS Maine; Reading Recovery Projects

ABSTRACT

Reading Recovery (RR), a one-on-one short-term intervention program for first-grade children at risk for literacy failure, targets the lowest 20% of a first-grade classroom. Currently, program guidelines specify that the kindergarten teacher recommend a list of children to be tested for the program. All recommended children are administered an assessment, the Observation Survey, which measures literacy skills. The children who score the lowest are taken into the program first, and the remaining children are placed on a waiting list. Some schools in Maine have adopted the practice of testing all entering first graders with the Observation Survey, as a beginning benchmark of progress for all children in the school system. A study addressed this issue quantitatively. Data were collected as part of the RR program in Maine during the 1995-96 school year. Two questions were explored: Is testing the entire class related to whether any at-risk children are ultimately not served by Reading Recovery? and Does testing the entire class more accurately delineate waiting list and RR children from their not-at-risk peers? Results indicate that although schools that test the whole class had a smaller (7% versus 19%) chance of failing to identify children in the fall school term who later went without needed services, the Chi square test did not reach statistical significance. On the whole, it appears that Maine kindergarten teachers do a good job of identifying the right children for testing. (NKA)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 463 526

Testing the Whole Class: What Impact Does It Have?

P.T Moore

1

U.S. DEPARTMENT OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Anne K. Rhodes-Kline
Center for Early Literacy
College of Education and Human Development
University of Maine
July, 1997

BEST COPY AVAILABLE

CS 510 851

Testing the Whole Class: What Impact Does it Have?

Reading Recovery is a one-on-one, short term intervention program for first grade children at risk for literacy failure. The program targets the lowest 20% of a first grade classroom. The decision regarding which children should be started into the program, and which ones should be started first, is not always a simple one. Currently, the program guidelines specify that the Kindergarten teacher recommends a list of children to be tested for the program. These are children who are behind the class on early literacy skills such as an interest in reading, knowing what to do with a book, and letter knowledge.

All of these recommended children are administered an assessment, the Observation Survey, which measures early literacy skills. The children who score the lowest are taken into the program first, and the remaining children are placed on a waiting list for the program. As the first Reading Recovery children are released from the program (which happens as soon as they are reading and writing at the average level of the other children in the classroom and are able to continue to learn reading and writing skills on their own), the lowest-scoring children from the waiting list can begin the program. If a child who is not initially identified as at risk begins to have difficulty later on, he or she can be added to the waiting list during the year.

For program evaluation purposes, the Observation Survey is administered to all children in Reading Recovery and to those on the waiting list for the program in both the fall and the spring of the first grade year. In addition, it is also administered to a random sample (generally 6 - 8 children per RR teacher) of the children not named on the Kindergarten teacher's list. The average scores of these children (called the "random sample") are used as a goal point for the skill levels of children in Reading Recovery.

Some schools in Maine have adopted the practice of testing all entering first graders with the Observation Survey, as a beginning benchmark of progress for all children in the school system. This practice operates as a safety net to prevent accidentally missing an at-risk child who did not appear to the Kindergarten teacher to be at risk. In smaller schools (those with fewer than 20 first graders, where it is not too much more work to test the whole class) this is the recommended practice. Some teachers at schools that test the entire entering first grade believe that this policy is the most reliable way to determine which children are at risk. These teachers sometimes report being surprised by low scores from children they thought were right on track.

The purpose of this study was to address this issue quantitatively. What impact does the policy of testing the entire entering first grade class have in terms of identifying at risk children accurately? Should all schools in Maine be encouraged to adopt this policy?

The data set available to answer the above question was collected as part of the Reading Recovery program in Maine for the 1995-96 school year. A number of variables were relevant to the present study. Schools were asked to indicate their policy for testing the incoming first grade class (i.e., whole class versus only those recommended and a random sample of the others). Among waiting list children, teachers were asked to indicate whether these children had been identified as at risk when they entered first grade (i.e., on the Kindergarten teacher's list) or later during the year. Observation Survey scores were obtained from children in all three groups (Reading Recovery, waiting list, random sample) in both the fall and the spring.

Two specific research questions were formulated based on the research problem and the data that were available. These were:

- Is testing the entire class related to whether any at-risk children are ultimately not served by Reading Recovery?
- Does testing the entire class more accurately delineate waiting list and RR children from their not-at-risk peers?

Is testing the entire class related to whether any at-risk children are ultimately not served by Reading Recovery?

Perhaps the biggest potential risk with not testing the whole class is that a truly at risk child will not receive Reading Recovery because his or her difficulty was not noticed by the Kindergarten teacher (assuming that it would have been detected on the Observation Survey.) In other words, the risk is that the Kindergarten teacher's assessment of which children are at risk of literacy failure is less valid than the Observation Survey's.

Method

A Chi square (χ^2) was conducted with the school as the unit of analysis. If a school had at least one child who was not identified at the beginning of the year as at risk, but was later added to the waiting list and was not served, the school was given a "yes" score. If the school had no children in this category, it was given a "no" score.

Results & Discussion

Although schools that test the whole class had a smaller (7% versus 19%) chance of failing to identify children in the fall who later went without needed services, the χ^2 test did not reach statistical significance ($\chi^2_1 = 3.43$, $p = .06$), nor was the effect very large ($r = .15$). Table 1 gives the counts for this analysis.

Table 1. Were Any Children Not Identified in the Fall Later Identified and Not Served?

School Policy	No		Yes		Total
Test Those on List and Random Sample of Others	89	(81%)	21	(19%)	110
Test Whole Class	40	(93%)	3	(7%)	43
Total	129		24		153

These are state data. On the whole, it appears that Maine Kindergarten teachers are doing a good job of identifying the right children for testing. However, in a school where the Kindergarten teacher is less aware of children's literacy tangles, or is too overburdened with other responsibilities to correctly identify all at risk children, the benefit of testing the entire entering first grade class is likely to be larger.

One limitation of the data to answer this question is that there is no measure of how many children were missed. A school either missed one or more at risk children, or it did not miss any. Another limitation of the data is that there was no way to know whether any children were initially not identified as at risk, but later identified *and served*. If a child was taken into the Reading Recovery program, information about when he or she was identified as at risk was not available about him or her. These limitations should be borne in mind when examining Table 1.

Does testing the entire class more accurately delineate waiting list and RR children from their not-at-risk peers?

Another way that the validity of the Kindergarten teachers' recommendations can be examined statewide is by looking at the overlap in fall scores between the children in the random sample (not at risk children) and the other two groups (waiting list and Reading Recovery children). The Reading Recovery program is designed to help out the minority of struggling

children in a classroom full of children who are mostly doing fine with literacy. It is generally assumed that the at-risk children's skills in the fall are far below those of their not-at-risk peers. In other words, when children are tested in the fall, there should be clear distinctions between the skills of at-risk children (those in the waiting list or Reading Recovery categories) and the skills of not-at-risk children (the random sample). Overlap between the scores of these groups might indicate that the children recommended for the program were not the lowest. Alternatively, it could indicate that the characteristics of the student population made the decision about which children should receive RR less clear (i.e., there were children in the "grey area").

Method

The second question was answered by examining this overlap with box-and-whisker plots, or box plots. Ten schools from each category (whole-class-tested or recommended-list) were randomly selected from the list of schools in the database. The box plots of each school on four assessments of literacy skill (Letter Identification, Concepts About Print, Hearing and Recording Sounds (HRS), and Writing Vocabulary) administered in the fall were examined. Overlaps between the fall scores of random sample and Reading Recovery, or between random sample and waiting list children were counted. These overlaps indicate the lack of clear delineation between at-risk and not-at-risk children. The four measures were chosen from the six tests of the Observation Survey because their distributions allowed the most discrimination among groups of entering first graders. The two measures not chosen suffer from "floor effects" because they assess skills most children do not possess at entry to first grade.

Results & Discussion

The box plots are given in the Appendix. Examination of the box plots reveals a lot of information about the schools and the students. For example, in the box plot of Fall Letter Identification, School number 13 and School number 16 have very different plots. School 16 had one Reading Recovery child score in the 20s (note the small circle below the box, denoting an "outlier"; extreme outliers are identified with asterisks), with the rest of the children from all three groups scoring in the 30s or higher. One conclusion might be that School 16 has a homogenous population of children; another explanation is that the Kindergarten teacher at School 16 does a very good job of teaching children their letters. In contrast to School 16, examine School 13. This school has a much broader range of letter identification skill among entering first graders. Notice that at least one Reading Recovery child scored in the low single digits, while the median, or middle score, for the random sample (note the bar running through the boxes - this bar denotes the median) is very close to the perfect score of 54.

Also note the numbers running along the bottom of the figure. These numbers, labelled with "N =" represent the number of children in each category within each school. School 11 had 8 random sample, 4 waiting list, and 8 Reading Recovery children, for example.

While much information can be gleaned from the boxplots, the detail that was important for the second research question concerned the overlaps between random sample children and the other two groups. An overlap occurs when the highest score for an at-risk child is higher than the lowest score for the random sample.

Table 2 gives a tally of the overlaps between the random sample and the other two groups. It is clear from Table 2 that there are no significant differences between schools that test

the whole class and those that rely on the recommendations of the Kindergarten teacher. For three of the assessment tasks, the numbers are identical. The other displayed non-significant differences in the opposite direction from what was expected.

Table 2. Overlaps Between the Scores of At-Risk Students and Not-At-Risk Students.

Literacy Assessment	Schools that use Recommended List Method (Schools 1 - 10)	Schools that Test Whole Class (Schools 11 - 20)
Letter Identification	10	10
Concepts About Print	10	10
Hearing and Recording Sounds (HRS)	8	8
Writing Vocabulary	9	10

An alternate explanation for the data in Table 2 is that there may have been waiting list children (in the "grey area") who caused the overlaps to occur. To address this possibility, overlaps were also counted between random sample and Reading Recovery groups only. Table 3 gives a tally of these overlaps. Again, there is no significant relationship between a school's testing policy and the number of overlaps between not-at-risk children's scores in the fall and those of children who would be served by Reading Recovery.

Table 3. Overlaps Between the Scores of Reading Recovery and Not-At-Risk Students.

Literacy Assessment	Schools that use Recommended List Method (Schools 1 - 10)	Schools that Test Whole Class (Schools 11 - 20)
Letter Identification	6	7
Concepts About Print	9	8
Hearing and Recording Sounds (HRS)	7	7
Writing Vocabulary	7	7

7

Conclusion

There is no evidence that Kindergarten teachers' judgements about who is at risk for literacy difficulties should be questioned on a statewide scale. Schools that test the entire entering first grade class do not differ on the measures chosen for this study from schools that test those recommended by the Kindergarten teacher and a random sample of the others. At this point, there is no reason to recommend that medium or large elementary schools with experienced and observant Kindergarten teachers test any more children with the Observation Survey than they are currently testing.

Testing the entire entering first grade class may still have an advantage for schools that wish to systematically track the progress of all students at both entry to and exit from first grade. Such practice can serve to establish benchmarks of progress (as recommended in Goals 2000) based on the skills of all first grade children in the school. Testing the entire class may also be the choice of schools that do not want to rely just on the judgement of a Kindergarten teacher, or in districts where not all entering first grade children have attended Kindergarten in the district.

The nature of the data available to answer this research inquiry limited the specific questions that could be posed. Perhaps the largest limitation is that, for question number one, only waiting list children could be included. So, if a school did not identify a child as at risk in the fall, but picked him or her up in Reading Recovery throughout the course of the year, even if that child did not finish the program, he or she would not have been counted. Future research into the possible benefits of testing the whole class should attempt to obtain this information.